Final Project: Web Scraping and Analysis

- **Introduction:**

In this project, you will extract data from a specific website, clean and analyze it using data processing techniques including Regular Expressions for text data handling. Then, you will visualize the findings and store the processed data in a database.

- **Project Details:**
  - **Data Extraction**
    - Identify the target website and collect the relevant data.
    - Store the extracted data in a structured file.
  - **Data Cleaning, Processing, and Regular Expressions**
    - Remove unnecessary or incorrect values.
    - Process textual data using **Regular Expressions (Regex)** to extract meaningful information (e.g., clean addresses, emails, or dates).
  - **Data Analysis**
    - Compute basic statistics from the extracted data.
    - Identify trends, patterns, and relationships between different variables.
  - **Data Visualization**
    - Create graphs and charts to present the findings clearly (e.g., bar charts, line graphs, or heatmaps).
  - **Data Storage**
    - Save the final processed data in a database (e.g., MongoDB) .

- **Project Deliverables:**
  - **Code** for data extraction
  - **File** containing raw extracted data
  - **Report** with analysis and findings
  - **Visual representations** of the data
  - **Database** with the final processed data

- **Bonus Task:**

Deploy your results using **Streamlit** to create an interactive web app for presenting the analysis.

Phase 1: Project Idea Submission and Discussion (Apr 15 th):

- Submit your project idea with a brief description of the website you'll be scraping data from and the type of data you'll be working with.
- Explain the overall goal of the project, how you plan to approach the data extraction, cleaning, analysis, and visualization process.
- This phase will be a submission and discussion of your project plan.

Phase 2:

- Perform all required tasks
- The deadline will be determined later.

Submission:
Phase_1→ https://forms.office.com/r/FRXDytcjXZ

Note:
This project will be carried out in teams, as registered in the previous form. Each team will work together to complete the project.

Good luck!